ANALELE ŞTIINȚIFICE ALE UNIVERSITĂȚII „ALEXANDRU IOAN CUZA" DIN IAŞI
Tomul LII/LIII Ştiințe Economice 2005/2006

POVERTY MAPPING OF ROMANIAN COUNTIES USING CLUSTER ANALYSIS

ALINA MĂRIUCA IONESCU*

Abstract

Poverty mapping plays a very important role in giving a visual representation of the intensity of poverty incidence by geographic area. This study tries to group Romanian counties taking into account several indicators that describe the various dimensions of poverty. The method used for grouping is cluster analysis.

The main objective of this paper is to show how poverty mapping using cluster analysis can be employed as a tool to identify homogenous poverty clusters of counties and help to reduce poverty by necessary resource allocation. Thus, the policy-makers can easily detect the most poverty affected areas and the poverty's specific in these areas and can be guided to target the poor for the best resource allocation to alleviate poverty.

Key words: poverty, poverty mapping, poverty alleviation, cluster analysis

1 Introduction

Poverty mapping is used for spatial identification of the poor (on which this paper concentrates) and serves to target poverty-alleviation programs from rural anti poverty programs to allocation of public services. It can assist practitioners in the formulation and implementation of poverty reduction, food security and sustainable development strategies, and in the monitoring of progress in poverty alleviation.

The choice of a poverty-mapping methodology depends on the objectives of the poverty mapping exercise, philosophical views on poverty, limits on data and analytical capacity and cost. Data availability is a fundamental constraint in choosing a poverty-mapping method. This constraint has two levels: the existence of data, and access to existing data. For example, using average values from disaggregated geographical units such as counties/districts (as it is used in this study) instead of household-unit data has the advantages that data requirements are less stringent and national statistical agencies may be more likely to release county/district level averages on request.

In this study, the method used for achieving poverty mapping is cluster analysis. It is applied in order to identify homogenous poverty clusters of counties aiming to reduce poverty by necessary resource allocation.

2 The method and the variables

"Cluster analysis is the art of finding groups in data." It aims "to form groups in such a way that objects in the same group are similar to each other, whereas objects in different groups are as dissimilar as possible" [Kaufman & Rousseeuw, 1990, 1]. The measurements

* PhD student, Department of Statistics, Faculty of Economy and Business Administration, University "Alexandru Ioan Cuza" Iaşi, e-mail: ali.ionescu@gmail.com

used in cluster analysis "can be organized in an n-by-p matrix, where the rows correspond to the objects (or cases) and the columns correspond to the variables" [Kaufman & Rousseeuw, 1990, 4]. In this example the cases are the n = 41 Romanian counties for which there are considered p=13 continuous variables describing poverty that cover almost all the dimensions of this phenomenon: persons employed in agriculture, as a percentage of the employed population – PAGR; persons living in rural areas, as a percentage of the total population – PRUR; poverty rate – RPOV; gross investments (thousands of millions – current prices) – GINV; life expectancy at birth (years) – LEXP; gross enrolment ratio – primary, secondary and tertiary schools (%) – GER; unemployment rate (%) – RUNP; infant mortality rate (per 1000 live births) – RINM; proportion of the population without sustainable access to electricity (%) – PWE; proportion of the population without sustainable access to water (%) – PWW; population per physician – P/PH; average net nominal monthly earnings – AVER; criminality rate (persons definitively convicted per 100000 inhabitants) – RCRI.

The Squared Euclidian Distance is chosen as proximity measure as it is frequently employed when working with interval data. Letting x_i and x_j represent two cases (counties) in the p-variate space (where p=13), the squared Euclidian distance between the two items x_i and x_j is defined as the sum of squared differences between the values for the items: $d_{ij}^2 = \sum_{f=1}^{13} (x_{if} - x_{jf})^2$.

The variables considered in this study are expressed in different measurement units: years, persons, thousands millions of ROL, percentage etc. Therefore, data values are standardized using z scores so as to equalize the effect of variables measured on different scales.

The reference year is 2002. For poverty rate the only available results at county level are for 2003. The data sources are UNDP Romania's National Human Development Report for 2001-2002 and CASPIS's (The Anti-Poverty and Social Inclusion Commission from Romania) statistics.

Statistical data processing was conducted using SPSS software.

As the city of Bucharest presents extreme values for some of the considered variables, it is not included in the study and needs to be investigated separately.

A principal components analysis was performed to verify if the chosen variables are relevant for this study. The high values of the extraction communalities show that all the variables fit well with the factor solution and could be kept in the analysis.

Due to the fact that the investigated population's size is relatively small (41counties), there are used hierarchical methods of clustering. In order to determine the most appropriate method for this study, there were applied all the seven hierarchical clustering methods available in SPSS. The resulting dendograms showed that Ward's method has differentiated the counties in the most clearly way and has found the most compact clusters.

3 Results of cluster analysis

3.1 Number of clusters

There is no exact procedure for determining the number of clusters. "To evaluate the number of clusters, one may always plot the criterion used to join clusters versus the number of clusters" [Timm, 2002, 534]. For example, a shape elbow in the plot of distances versus the number of clusters may be an indication of the number of clusters.

Table no. 1 *– Agglomeration Schedule*

Stage	Cluster Combined		Coefficients	Stage Cluster First		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	26	1.214	0	0	21
⋮						
30	1	2	157.402	17	24	33
31	16	41	173.065	19	0	35
32	9	28	**190.769**	25	26	37
33	1	11	**208.875**	30	23	36
34	7	10	227.734	21	29	35
35	7	16	**249.398**	34	31	38
36	1	6	**271.881**	33	0	39
37	3	9	294.920	28	32	40
38	7	25	**333.013**	35	27	39
39	1	7	**390.379**	36	38	40
40	1	3	520.000	39	37	0

Source: Results obtained with SPSS

Table no. 1 shows how the counties are clustered together at each stage of the cluster analysis. The Coefficients column indicates the distance between the two clusters (or cases) joined at each stage. The values here depend on the proximity measure (Squared Euclidian Distance) and linkage method (Ward's method) used in the analysis.

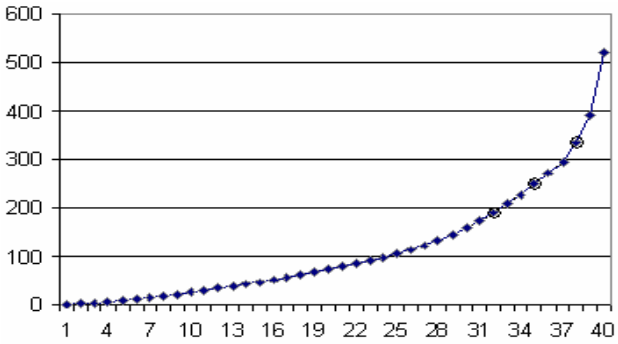

Fig.1 The agglomeration schedule coefficients plot

For a good cluster solution, a sudden jump can be seen in the distance coefficient as it can be read down the table. The stage before the sudden change indicates the optimal stopping point for merging clusters. For this example, we should consider using a 9, 6 or 3-cluster solution. For a better visualization of this criterion the "hockey stick" plot of agglomeration schedule coefficients is displayed in figure 1. It can be easily seen that 9 clusters remain after stage 32, 6 clusters after stage 35 and 3 clusters after stage 38.

These three solutions are also illustrated in figure 2 which presents the dendogram.

At a first look the dendogram shows three obvious clusters, which can be interpreted as it follows: one cluster consists of the most developed counties (AG, PH, MS, AB, DJ, IS, CT, HD, BV, SB, TM, CJ), another cluster refers to the counties with moderate intensity of poverty (BR, CS, CV, HR, GL, GJ, DB, OT, VL, BN, GR, SJ, IF, AR, BH, MM, SM) and the other cluster groups the most poverty affected counties (CL, IL, TR, TL, BC, MH, NT, BZ, SV, VN, BT, VS). The counties groupings in 6 or in 9 clusters are the solutions that

differentiate the most clearly the clusters and identify the special case of Vaslui County that forms a cluster by itself.

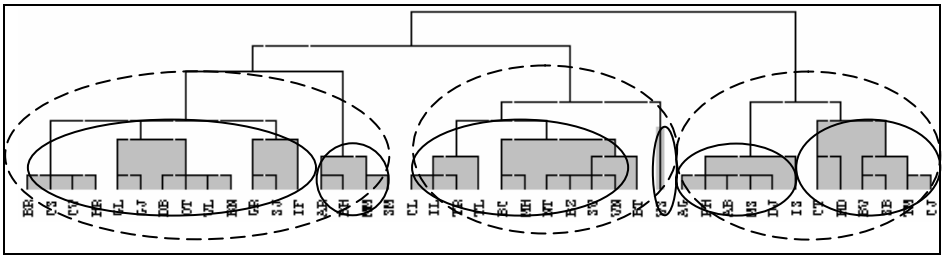

Fig.2 Ward's dendogram for the 3, 6 and 9 clusters solution

3.2 Territorial distribution of the clusters

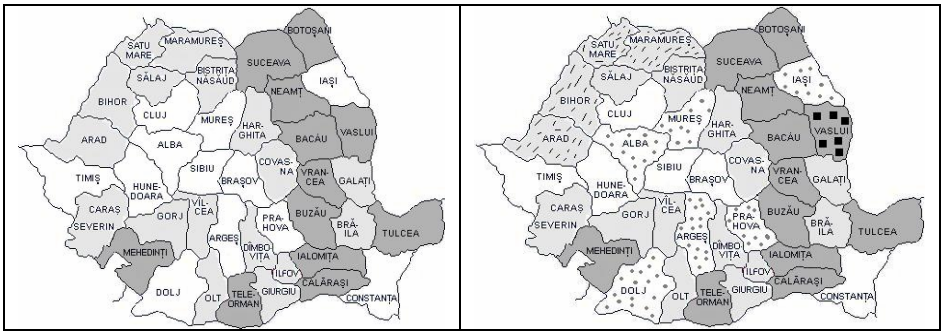

Fig. 3 Territorial distributions of the clusters for the 3 and 6 cluster solutions

The 9 clusters – solution reproduces well enough the geographic repartition of Romanian counties as it groups by twos and threes neighbor counties. This solution is efficient when preparing anti-poverty programs and policies that are to be applied to small areas.

If one wants to develop programs that focus on large areas (like regions of a country) then it is recommended to use the 6 clusters – classification as it reproduces more clearly the geographical map of counties.

The 3 clusters-solution provides the possibility of identifying three main directions for poverty alleviation programs and policies: monitoring activities for the cluster that consists of the most developed counties, specific measures for improvement of socio-economic indicators for the cluster of counties with moderate intensity of poverty and allocation of important resources and implementation of radical programs for the cluster of the most poverty affected counties.

3.3 Clusters' profiles

Once the clusters are obtained, it is generally useful to describe each group using some descriptive tools to create a better understanding of the differences that exist among the created groups. In order to characterize the clusters, there are computed descriptive statistics (means) for each cluster (table no. 2).

Table no. 2 – Descriptive statistics (means) for each of the 6 clusters

Clust.	1	2	3	4	5	6
PAGR	51.46	36.14	56.51	42.71	25.44	39.83
PRUR	59.80	49.75	59.90	58.05	30.70	50.80
RPOV	0.37	0.29	0.40	0.32	0.23	0.27
GINV	4129.27	11632.83	1497.00	5100.00	16271.17	5800.00
LEXP	70.79	71.32	71.10	71.34	71.20	69.48
GER	59.04	70.20	58.00	60.64	75.03	63.78
RUNP	9.80	8.47	15.90	9.44	8.60	4.68
RINM	20.16	18.63	23.20	13.82	17.72	18.08
PWE	2.58	2.32	6.80	2.07	1.50	2.40
PWW	39.16	30.57	53.50	34.12	11.20	21.85
P/PH	832.36	464.83	971.00	709.54	397.50	528.25
AVER	3391392.91	3622089.00	33.6616.00	3668304.31	3793331.17	3231682.50
RCRI	459.64	319.83	547.00	347.15	322.17	419.25

Source: Results obtained with SPSS

The highest level of poverty is registered by cluster 3, formed of Vaslui County that is extremely affected by unemployment; in 2002 unemployment rate was 15.9%. Population sustainable access to electricity and water (essential elements for a decent life), is very low: 6.8% (a great percentage comparing to other clusters) of population doesn't have sustainable access to electricity and more than a half of this county population doesn't have access to sustainable water. These facts and the very small number of physicians in this area could be the causes for the highest infant mortality rate. The population of Vaslui County is the most poverty affected in almost all the dimensions of this phenomenon: the lowest gross enrolment ratio, the lowest access to health services, to water and electricity, the highest criminality rate. The economy of Vaslui County is in crisis as it presents the lowest level of gross investments (3 times lower than the next cluster and 11 times lower than the richest cluster). This could be a justification of the very high unemployment rate together with the highest proportion of persons occupied in agriculture (over 56% of the total employed population) and of persons living in rural areas (almost 60% of the total population).

Another especially poverty affected cluster is cluster 1 (BC, MH, NT, BZ, SV, VN BT) which presents alarming values for all the variables.

Cluster 5, that includes CT, HD, BV, SB, TM, CJ, presents the highest standard of living as it is characterized by: a very low percentage of rural population (30.7%) and of population employed in agriculture (25.44%) comparing to other clusters, the lowest poverty rate, the highest level of gross investments, the highest gross enrolment ratio, the lowest unemployment rate and infant mortality rate, the best population access to health services and to utilities (water and electricity). The high standard of living is also reflected by one of the lowest criminality rate.

Another group of counties with low level of poverty is cluster 2 (AG, PH, AB, MS, DJ, IS) that presents closer values to those of cluster 5 for most of the considered variables.

4 Conclusions

Cluster analysis permitted to group the 41 counties of Romania in homogenous groups considering the poverty dimensions such as: health, education, unemployment.

According to each cluster's profile there could be designed and developed specific poverty alleviation programs that take into account poverty intensity in each considered dimension. Therefore, to the clusters that present deprivations in health dimension of

poverty should be designed and applied appropriate programs to improve the access to health services. For clusters with low level of investments and high unemployment rates there can be allocated resources to stimulate investments so as to create new jobs and reduce unemployment. The programs that focus on infrastructure development can target the groups of counties characterized by low sustainable access to electricity and water correlated with high percentage of rural population and population employed in agriculture.

In conclusion, cluster analysis employed in poverty mapping may be of a real utility in designing poverty reduction programs and policies as it permits to detect the most poverty affected areas and the poverty's specific in these areas and help the policy-makers to target the poor for the best resource allocation to alleviate poverty.

Bibliography

Anderberg, M. R., *Cluster Analysis for applications*, Academic Press, New York, 1973
Davis, B., *Choosing a method for poverty mapping*, 2003, la http://www.povertymap.net/publications/doc/CMPM%20DAVIS%2013%20apr03%20sec.pdf, accessed on 3 January 2006.
Everitt, B., Landau, S., Leese, M., *Cluster analysis,* 4th Edition, Edward Arnold Publishers Ltd., London, 2001.
Garson, G. D., *Quantitative Research in Public Administration, PA 765 Statnotes: An Online Textbook*, 2005, la http://www2.chass.ncsu.edu/garson/pa765/cluster.htm, accessed on 5 December 2005.
Jaba, E., *Statistica*, Ediția a treia, Editura Economică, Bucureşti, 2002.
Kaufman, L. and Rousseeuw, P. J., *Finding groups in data: An introduction to cluster analysis*, John Wiley & Sons, New York, 1990.
Timm, N., *Applied Multivariate Analysis*, Springer Text in Statistics, 2002.
*** www.caspis.ro, accessed on 10 December 2005.
*** www.spss.com, accessed on 12 November 2005.
*** www.undp.ro, accessed on 20 November 2005.